REPORT DOCUMENTATION PAGE

AFRL-SR-BL-TR-01-

0335

1. REPORT DATE *(DD-MM-YYYY)* 09-01-2001	**2. REPORT TYPE** Final	**3. DATES COVERED** *(From - To)* Mar 1, 1997 - Aug 31, 2000
4. TITLE AND SUBTITLE INFORMATION DYNAMICS IN IMAGE DECONVOLUTION		**5a. CONTRACT NUMBER** F49620-97-1-0148
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Sudhakar Prasad		**5d. PROJECT NUMBER**
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) University of New Mexico, Center for Advanced Studies, 800 Yale Blvd. NE, Albuquerque, NM 87131		20010612 029
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Office of Scientific Research, Directorate of Physics and Electronics		**10. SPONSOR/MONITOR'S ACRONYM(S)** AFOSR/NE
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Unlimited

13 SUPPLEMENTARY NOTES

14. ABSTRACT This report consists of somewhat abridged versions of all previously submitted progress reports, placed here in individual sections, followed by a section on the progress made in the last 12-month performance period and on the promising future directions for our work. A broad understanding of all aspects of image formation, detection, and processing in information theoretic terms has been achieved in our studies under direct support of the Contract. Specifically, we have (i) shown how prior knowledge about the object being imaged, particularly about the support of its intensity distribution, can be regarded as a gain of information; (ii) computed the theoretical limit on information throughput, i.e., the channel capacity of an astronomical imaging system; (iii) computed theoretically and in simulation the detailed characteristics of the dynamics of noise versus information in imaging under turbulence when support information is employed as a post processing tool; and (iv) gained a preliminary understanding of the inverse relation between noise and information.

15. SUBJECT TERMS
Statistical information, image processing, deconvolution, imaging under turbulence

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Sudhakar Prasad
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU	21	**19b. TELEPHONE NUMBER** *(include area code)* 505-277-5876

ABSTRACT

This final report on the Grant No. F49620-97-1-0148 consists of somewhat abridged versions of all previously submitted progress reports, placed in individual sections, followed by a section on the progress made in the last 12-month performance period and on promising future directions for our work.

A broad understanding of all aspects of image formation, detection, and processing in information theoretic terms has been achieved in our studies under direct support of the grant. Specifically, we have (i) shown how prior knowledge about the object being imaged, particularly about the support of its intensity distribution, can be regarded as a gain of information; (ii) computed the theoretical limit on the information throughput, i.e., the channel capacity, of an astronomical imaging system; (iii) computed theoretically and in simulation the detailed characteristics of the dynamics of noise versus information in imaging under turbulence when support information is employed as a post-processing tool; and (iv) gained a preliminary understanding of the inverse relation between noise and information.

I. REPORT ON PROGRESS MADE DURING THE PERIOD 3/97-8/97

The first recognition that image formation may be usefully described from the viewpoint of Shannon's statistical information theory dates back to Fellgett and Linfoot's pioneering work [1] in 1955. They argued for two competing viewpoints on the assessment of the quality of image formation:

1. How similar is the image to the object? This could be studied by defining a metric, that measures the departure of the image from the object, in terms of optical system transmission factors.

2. How faithfully does the system image a class of objects, in the sense of the closeness of the information content of the image set relative to that of the object set?

By shifting attention from a single object to a whole class of objects, the second viewpoint provides for a very powerful characterization of the imaging system as a conduit for the passage of information. This is an inherently more useful viewpoint if the objective of study is the optimality of the design of an optical system to image a variety of objects. For example, if a whole class of satellites which all contain the same gross features is the set of objects for which an imaging system is being designed, then it is clear that the second viewpoint provides a statistical criterion for the best design. Indeed, since one does not typically know in advance the details of the object being imaged, the question of how close its image is to itself cannot even be answered.

It is this motivation that characterizes our work, based on the use of statistical information entropy and other related ideas.

A. Object Set

We can specify the object set statistically by a probability distribution $P[\tilde{O}(\vec{f})]$ on the object spatial spectrum $\tilde{O}(\vec{f})$, the latter defined as the Fourier transform of the two-dimensional intensity distribution of the object. The object illumination is assumed incoherent here, as is true for most astronomical sources or solar-illuminated satellites. The object spatial spectrum at a spatial frequency is directly proportional to the complex fringe visibility for interference of light received at two apertures with vector separation proportional to the spatial frequency vector. This is the essence of the Van Cittert – Zernike theorem, which governs interferometric image reconstruction.

We begin with a circular Gaussian model for $P[\tilde{O}(\vec{f})]$:

$$P[\tilde{O}(\vec{f})] = \prod_{\vec{f}} \frac{1}{2\pi\sigma_0^2(\vec{f})} e^{-|\tilde{O}(\vec{f})-\langle\tilde{O}(\vec{f})\rangle|^2/(2\sigma_0^2(\vec{f}))}, \tag{1}$$

where the product is over a discrete set of spatial frequencies and $\sigma_0(\vec{f})$represents the

standard deviation of the object spectrum at frequency $\vec{f}$. The question we asked and successfully answered is the following: Relative to this object set, what is the statistical information successfully transmitted by a simple optical system working under conditions of turbulence induced phase aberrations at high light levels? In order to answer the question, we must first discuss how an optical system processes an image under turbulent conditions, but without the complications of either the detector or low-light-level shot noise.

B. Imaging under Turbulent Conditions

An incoherent imaging system is a linear system in the sense that the image of a collection of incoherent point sources is a superposition of the elementary images (point spread functions (PSFs)) of the individual point sources. Since an incoherent object may be regarded as a contiguous collection of such point sources, the image and object intensity distributions may shown to be related through the PSF by a convolution. In the Fourier space, therefore, the object and image spatial-spectrum amplitudes, $\tilde{I}(\vec{f})$ and $\tilde{O}(\vec{f})$, are related by a simple product involving the Fourier transform of the PSF, also known as the optical transfer function (OTF), $\mathcal{H}(\vec{f})$:

$$\tilde{I}(\vec{f}) = \tilde{O}(\vec{f})\mathcal{H}(\vec{f}). \tag{2}$$

The OTF of an optical system under turbulent conditions is on average a product of the system OTF in the absence of turbulence and the OTF that can be directly associated with the turbulent-induced amplitude and phase fluctuations.

Of course, under turbulence, there arise fluctuations in the recorded image intensity spectrum, since the OTF suffers from fluctuations. We may see from Eq. (2) that the image noise is multiplicative, which is a major analytical complication. We adopt an approximate procedure to include this noise in the analytical treatment that follows.

As we have indicated, the mean OTF is a product of the system OTF $\mathcal{H}_0(\vec{f})$ and the mean atmospheric OTF $\langle\mathcal{H}_{atm}(\vec{f})\rangle$:

$$\langle\mathcal{H}(\vec{f})\rangle = \mathcal{H}_0(\vec{f})\langle\mathcal{H}_{atm}(\vec{f})\rangle. \tag{3}$$

The sum of the variances of the real and imaginary parts of the OTF, which we simply call the psuedovariance of the OTF, may be approximately computed with a slightly modified structure function for the turbulence-induced random phase distribution in the pupil. The result is

$$\mathrm{var}[\mathcal{H}(\vec{f})] \approx [1 - \exp(-C\lambda^2F^2|\vec{f}|^2/r_0^2)]|\mathcal{H}_0(\vec{f})|^2, \tag{4}$$

where F is the system focal length, λ the light wavelength, r_0 the atmospheric coherence diameter, and C is a constant of order 3.44. The correlation of the fluctuations of the OTF at two different spatial frequencies may also be computed analytically with the result,

$$\langle\Delta\mathcal{H}(\vec{f})\Delta\mathcal{H}^*(\vec{f}')\rangle \approx \mathrm{var}[\mathcal{H}(\vec{f})]\delta_{\vec{f},\vec{f}'}, \tag{5}$$

under the assumption of strong turbulence $r_0 \rightarrow 0$. In other words, in this limit, two different spatial frequencies have uncorrelated noise.

It is worth noting that the preceding analysis is a conditional analysis of the mean and the fluctuations in the image, conditioned on the precise specification of the object. We must bear in mind that an object is drawn from a statistical set, the so-called object set. But it is the conditional statistical distribution of this sort that governs the degradation of information in an object by the random fluctuations of the processing system. When the related conditional information is averaged over the object set, we get a measure of what portion of the information in the image is unreliable and therefore not faithfully transmitted from the object set. Shannon defined the successfully transmitted information as the following mutual information:

$$H[O;I] = H[I] - H[I/O] = H[O] - H[O/I], \tag{6}$$

where the object information content is given by

$$H[O] = -\sum_{\tilde{O}} P[\tilde{O}] \log P[\tilde{O}], \tag{7}$$

with the object probability distribution given by expression (1). Similar expression may be given for the image information $H[I]$. For the conditional information $H[I/O]$ (or $H[O/I]$), one has a similar expression:

$$H[I/O] = -\sum_{\tilde{O},\tilde{I}} P[\tilde{O}] P[\tilde{I}/\tilde{O}] \log P[\tilde{I}/\tilde{O}], \tag{8}$$

where $P[\tilde{I}/\tilde{O}]$ is the conditional probability distribution of the image spectrum, given a specified object spectrum. If we make the tenable assumption that the OTF $\mathcal{H}(\vec{f})$ in Eq. (2) is a result of a linear superposition of many independent contributions, as true in the strong-turbulence limit, then the conditional statistics of the image spectrum can be accurately described by a circular Gaussian probability distribution with the mean and the variance given by Eqs. (3) and (4).

With the preceding justifications and approximations, we can now discuss two cases of some practical interest.

1. Low Variability in the Object Set, $\sigma_0(\vec{f}) \rightarrow 0$

In this case, the image spectrum probability distribution, which requires an integration of the object probability distribution with the conditional distribution $P[\tilde{I}/\tilde{O}]$ over the object realizations, reduces approximately to a product of two Gaussian distributions. Such integrations are straightforward. The final result for the *successfully transmitted* information, the mutual information of Eq. (6), is the following:

$$H[O;I] = \sum_{\vec{f}} \log\left[1 + \sigma_0^2(\vec{f})|\langle\mathcal{H}(\vec{f})\rangle|^2/\sigma_{I/O}^2(\vec{f})\right],$$
$$\to A \int d^2 f \log\left[1 + \sigma_0^2(\vec{f})|\langle\mathcal{H}(\vec{f})\rangle|^2/\sigma_{I/O}^2(\vec{f})\right], \tag{9}$$

where the conditional variance $\sigma_{I/O}^2(\vec{f})$ in the image when the object is known *a priori* takes the form

$$\sigma_{I/O}^2 \approx |\langle\tilde{O}(\vec{f})\rangle|^2 |\mathcal{H}_0(\vec{f})|^2. \tag{10}$$

There are three remarks in order here. Note that frequencies beyond the system cutoff are excluded from Eq. (9), since for such frequencies no information is transmitted by the system. When the conditional variance in the image $\sigma_{I/O}$ vanishes, the system transmitts perfectly the full, infinite amount of information there is in each of the infinitely accurately specified objects. (This can be made finite by requiring only a certain finite decimal accuracy of the object spectrum specification.) Finally, when the object variability goes away, $\sigma_0(\vec{f}) \to 0$, then there is no information to transmit. A fixed object carries no information in the sense of its potential to carry a signal.

2. Infinitely Strong Turbulence, $r_0 \to 0$

In this limit, the information successfully transmitted by the optical system becomes independent of the object set information, in particular the variance of the object set. This is understandable, since in the strong-turbulence limit the atmosphere filters out all but the lowest spatial frequencies, and so it is expected that object variability will have little bearing on the information actually passed on to the image.

II. REPORT ON PROGRESS MADE DURING THE PERIOD 9/97–8/98

The work during this period was focused on gaining the vital understanding of how an *a priori* constraint like the knowledge of the support of the image, to be simply called the support constraint hereonwards, can lead to improved statistical information in the image. An indirect, although somewhat more traditional, approach that we first took was to evaluate the reduction of noise brought about by the imposition of the support constraint. We summarize below only the salient features of this work, but leave its more complete discussion to the refereed publication [3] accompanying this Report. The second part of the report is a discussion of our more direct approach to the problem, in which we discuss the information content of the image before and after the imposition of the support constraint.

A. Noise Reduction in Image Deconvolution with Support Constraint

Use of *a priori* constraints is often made in image reconstruction algorithms to improve the final image beyond what the limited observational data can generate. One way in which the improvement can take place is through noise reduction that is directly enabled by the *a priori* information. We summarize here how the knowledge of the support of the object can lead to significant noise reduction in a linear deconvolution of raw images that are degraded by atmospheric turbulence.

The spatial spectra of the object and the image, $O(\vec{u})$ and $I(\vec{u})$ are related very simply through the optical transfer function (OTF) $H(\vec{u})$,

$$I(\vec{u}) = O(\vec{u})H(\vec{u}). \tag{1}$$

For well-corrected optics and at high light levels with noiseless detectors, the image noise in the spatial frequency ($\vec{u}$) plane is governed by the fluctuations of $H(\vec{u})$ arising from turbulence.

When the image is constrained to have a known finite support described by the support function $s(\vec{x})$, equal to 1 for $\vec{x}$ inside the support and 0 outside, then the support-constrained and unconstrained images are related by the multiplicative factor $s(\vec{x})$ in the physical domain and by convolution with the support spectrum $S(\vec{u})$ in the Fourier plane:

$$I_s(\vec{u}) = \int I(\vec{u} - \vec{u}')S(\vec{u}')d^2u'. \tag{2}$$

The spatial spectrum (simply called spectrum hereafter) of the image noise can be constructed in terms of the two mean squares $N(\vec{u}) \equiv \langle |I(\vec{u}) - \langle I(\vec{u})\rangle|^2\rangle$ and $M(\vec{u}) \equiv \langle [I(\vec{u}) - \langle I(\vec{u})\rangle]^2\rangle$. The relation between the noise power spectra before and after imposition of the support condition follows from Eq. (2) supplemented by the result (1). The following approximate convolution relation may be shown [3] to hold for the mean squared modulus:

$$N_s(\vec{u}) = \int d^2u_1 N(\vec{u} - \vec{u}_1)S(\vec{u}_1)e^{-(1/2)D_\phi(\lambda f u_1)}, \tag{3}$$

where D_ϕ represents the phase structure function for atmospheric turbulence.

A powerful, although somewhat approximate, picture of the relation (3) between the constrained and unconstrained noise spectra emerges when one expands $N(\vec{u} - \vec{u}_1)$ in a Taylor series around $\vec{u}_1 = 0$. Such an expansion leads to an expansion of Eq. (3) in derivatives of $N(\vec{u})$ with increasing derivative order from one term to the next. The coefficients of the derivative terms are expressed as tensor moments of the support spectrum. With the retention of only the first three terms in the expansion, we obtain

$$\begin{aligned}\Delta N(\vec{u}) &\equiv N_s(\vec{u}) - N(\vec{u}) \\ &= \vec{B} \cdot \vec{\nabla} N(\vec{u}) + \frac{1}{2}\hat{\mathbf{D}} : \vec{\nabla}\vec{\nabla} N(\vec{u}),\end{aligned} \tag{4}$$

where

$$\vec{B} = -\int d^2u \; S(\vec{u}) e^{-(1/2)D_\phi(\lambda f u)} \vec{u} \tag{5}$$

is a vector and

$$\mathbf{D} = \int d^2u \; S(\vec{u}) e^{-(1/2)D_\phi(\lambda f u)} \vec{u}\vec{u} \tag{6}$$

is a second-rank tensor.

Note that the equation (4) formally describes a noise transformation that contains a drift, or ballistic transport, term and a diffusive spreading term. The physical implications of this viewpoint are noted in detail in Ref. 3.

For an inversion-symmetric support like a circle (or an ellipse), the drift vector $\vec{B}$ vanishes by symmetry. The noise transport induced by the support constraint in the spatial frequency plane is then purely diffusive. Such transport enables a reduction of noise. All one has to do is to leave unchanged the original, unconstrained data at those spatial frequency locations where the transformed data are noisier than the original data, as pointed out by Tyler and Matson [4] in their simulations of the same problem. With this picture, we can easily evaluate the total noise reduction possible in one application of the support constraint. It turns out, as shown in detail in the attached manuscript, that we have an analytical expression for the ratio of the total noise reduction to the total noise in the original image:

$$F^{(d)}(\sigma) \equiv \frac{\Delta W^{(d)}}{W^{(d)}} \approx \left[\frac{\pi\sigma}{\lambda f / r_0}\right]^2 e^{-\frac{\pi^2\sigma^2}{2C(\lambda f/r_0)^2}}, \tag{7}$$

where σ is the diameter of the support, $\lambda f / r_0$ is the diameter of the atmospheric seeing disk, and C is a number of order 10.

Note from Eq. (7) that the best fractional noise reduction $F^{(d)}$ is obtained when the support is comparable in size to the seeing disk. Equation (7) represents one of the most important results of the work, and is confirmed, more or less quantitatively, by the simulations of Tyler and Matson [4]. In the next subsection, we begin a description of this problem from the viewpoint of information theory. Details may be found in another accompanying article [5].

B. Information Dynamics in Seeing-Limited Image Deconvolution with Support Constraint

Consider the mutual information, as defined by Shannon to be the information successfully transmitted through a communication channel. The mutual information $\mathcal{I}(\vec{u})$ at spatial frequency $\vec{u}$ of the image spectrum is the difference,

$$\mathcal{I}(\vec{u}) = \mathcal{H}_o(\vec{u}) - \mathcal{H}_{o|i}(\vec{u}), \tag{8}$$

of the (unconditional) object information $\mathcal{H}_o(\vec{u})$ and the conditional object information $\mathcal{H}_{o|i}(\vec{u})$, conditioned on the observation of the image spectrum. According to Shannon, each kind of information is given by averaging the logarithm of the corresponding probability density function (pdf):

$$\mathcal{H}_o(\vec{u}) = -\langle \log P_o(O(\vec{u})) \rangle = -\int P_o(O(\vec{u})) \log P_o(O(\vec{u})) d^2 O(\vec{u});$$

$$\mathcal{H}_{o|i}(\vec{u}) = -\langle \log P_{o|i}(O(\vec{u})|I(\vec{u})) \rangle = -\int P(O(\vec{u}), I(\vec{u})) \log P_{o|i}(O(\vec{u})|I(\vec{u})) d^2 O(\vec{u}) d^2 I(\vec{u}). \tag{9}$$

The actual values that the statistical variables o, i can assume are always denoted by the corresponding upper-case symbols O, I.

The pdf's are related because of the imaging equation (1). For a turbulent atmosphere, as true for a ground-based imaging system, the OTF H has random variations that can be described adequately by the Kolmogoroff hypothesis [6].

An application of the support constraint in the image domain changes the image spectrum by means of a convolution of the original image spectrum and the spectrum of the support function. Since the object spectrum is unchanged, the application of the support constraint causes a change in the system's mutual information (8) by amount

$$\Delta\mathcal{I}(\vec{u}) = -\left[\mathcal{H}^{(s)}_{o|i}(\vec{u}) - \mathcal{H}_{o|i}(\vec{u})\right]. \tag{10}$$

It is enough for our purposes thus to compute the change in the negative conditional information $-\mathcal{H}_{o|i}$.

Assuming that the change of pdf's under the support constraint is small, we can carry out a first-order analysis for the change $\Delta\mathcal{I}$. After some tedious algebra, one may show that correct to the lowest-order changes of the pdf's, the change in information throughput is given by the simple relation

$$\Delta\mathcal{I}(\vec{u}) = \int\int \left[\Delta P_{i|o}(I(\vec{u})|O(\vec{u}))\right] P_o(O(\vec{u})) \log P_{o|i}(O(\vec{u})|I(\vec{u})) d^2 O(\vec{u}) d^2 I(\vec{u}). \tag{11}$$

The conditional pdf $P_{i|o}$, when the support constraint is not applied yet, is easily evaluated in terms of the known object pdf and the pdf of the OTF which we can compute based on the Kolmogoroff hypothesis of turbulence. Indeed from Eq. (1), since for a given $O(\vec{u})$, $I(\vec{u})$ and $H(\vec{u})$ are proportional to each other, the conditional pdf $P_{i|o}$ is simply related to the pdf P_H of the OTF:

$$P_{i|o}(I(\vec{u})|O(\vec{u})) = P_h(H(\vec{u}) = I(\vec{u})/O(\vec{u})) J^{2)}(H/I), \tag{12}$$

where $J^{(2)}(H(\vec{u})/I(\vec{u}))$ is the Jacobian determinant of the transformation from the complex $H(\vec{u})$-plane to the complex $I(\vec{u})$-plane. Since $O(\vec{u})$ is held fixed in the conditional pdf

calculation, the Jacobian is simple and turns out to have the value $1/|O(\vec{u})|^2$. The inverse conditional pdf $P_{o|i}$ can in turn be computed by means of the Bayes theorem,

$$P_{o|i}(O(\vec{u})|I(\vec{u})) = \frac{P_{i|o}(I(\vec{u})|O(\vec{u}))P_o(O(\vec{u}))}{\int P_{i|o}(I(\vec{u})|O(\vec{u}))P_o(O(\vec{u}))d^2O(\vec{u})}. \quad (13)$$

When other sources of noise are present as well, as for example in photon counting at low count rates, their effects may once again be integrated into a single pdf by further applications of the Bayes theorem for compounded random processes.

When the support constraint is applied in the image, the image spectrum changes via Eq. (2). However, because of the linearity of that equation, the statistics of the post-support spectrum $I^{(s)}(\vec{u})$ are rather simply related to those of the pre-support spectrum $I(\vec{u})$. It may be shwon under rather general conditions that the statistics of the OTF $H(\vec{u})$ are Gaussian and therefore, by Eq. (12), so are the conditional statistics of i. Since a linear combination of Gaussian random variates too has Gaussian statistics, the conditional statistics of the post-support spectrum are also Gaussian. Under these conditions, the change of mutual information $\Delta\mathcal{I}(\vec{u})$ given by Eq. (11) takes on a more explicit final form

$$\Delta\mathcal{I}(\vec{u}) = -\frac{1}{\sigma_h^2}\Delta\left\langle|\frac{I(\vec{u})}{O(\vec{u})} - \langle H(\vec{u})\rangle|^2\right\rangle - \int \Delta P_i(I(\vec{u}))\log P_i(I(\vec{u}))d^2I(\vec{u}), \quad (14)$$

where σ_h^2 denotes the variance of the circular Gaussian pdf which describes the fluctuations of the OTF. Like in Eq. (11), the prefix Δ designates the change in a quantity when the support constraint is applied.

III. REPORT ON PROGRESS MADE DURING THE PERIOD 9/98–8/99

Three major accomplishments were made during this period. The first concerns the publication of the paper "Dynamics of turbulence induced noise in image deconvolution with support constraint" in the J. of the Optical Society of America (JOSA A, vol. 16, pp. 1769-1778 (1999)). In this paper, discussed in the Sec. II A and accompanying the report, we discuss the improvement of the information content of the decovolved image in somewhat indirect terms of reduction of noise in the image when *a priori* knowledge of the image support is used. The second concerns a direct study of the information transport under the application of support constraint. This paper, discussed in Sec. II B and attached here in reprint form, was presented at the *Fundamental Issues in Image Formation, Detection, and Processing* workshop organized by me at UNM's Center for Advanced Studies, February 6-7, 1999. Numerical simulations confirming the improvement of mutual information are presented here in detail. The third accomplishment is our explicit calculation of the information capacity of an imaging system limited by atmospheric turbulence and additive detector noise. This work [7], attached here in preprint form, was invited, presented, and well received at the recently concluded Air Force Maui Optical Station (AMOS) Technical Conference, Aug 30 – Sep 3, 1999. A longer and more complete paper [8] on this work

was submitted, refereed, and published in Optics Communications. This work has made it possible for us to begin to develop a quantitative understanding of the statistical information content of *a priori* knowledge such as support, which was a key task that we had set out to perform under the present Contract. Finally, a paper entitled "Noise Transport and Removal in a Speckle Imaging Algorithm," co-authored with David Tyler, was presented at the Annual Optical Society of America Meeting in Santa Clara, September 26-30, 1999.

A. Preliminary Simulations of Information Dynamics

We have used a variety of computer programs and a telescope imaging simulation code to generate ensembles of data and corresponding mutual information plots. First, an object ensemble of five different scenes was synthesized using Gaussian "stars" with various widths and relative intensities. These objects were used as inputs to a simulation modelling the effects of the turbulent atmosphere, telescope diffraction, and finite-resolution detection. The simulation generated ensembles of several hundred frames of noise-corrupted imagery, all of which were then processed using a deconvolution algorithm (to increase resolution) and our support-constraint convex-projections algorithm (to remove noise). Finally, the mutual information over the data and object ensembles was calculated; first for data processed with only the deconvolution algorithm and then for data processed with both algorithms. The mutual information was calculated for each of the central 32×32 pixels in the image arrays, forming information "maps" of the image field. The displays (Figs. 1, 2, and 3) presented here show the spatial distribution of the statistical information in the object set, mutual information at high turbulence ($D/r_0 = 10$), and the change of mutual information in the image when the support constraint is applied. The overall improvement of MI by about 5-10% is consistent with the degree of equivalent noise reduction under the same support application.

B. Information Capacity of a Seeing Limited Imaging System

In order to gain fundamental insights into the nature of information transport and improvement when *a priori* constraints are applied, we must first understand the maximum information carrying capacity of an imaging system.

We generalize our treatment of Sec. II to include an additive, zero-mean, Gaussian random noise source,

$$I(\vec{u}) = O(\vec{u})H(\vec{u}) + E(\vec{u}), \tag{1}$$

and to include at once a finite region of the spatial frequency plane in our considerations. The description at the level of individual pixels is not complete because it ignores interpixel correlations over a finite region. The definitions must be suitably modified, with functional integrals replacing ordinary integrals on the right-hand sides.

For $D/r_0 >> 1$, as before, the OTF $[H(\vec{u})]$ has Gaussian statistics, under which the

conditional statistics of the image spectrum $[I(\vec{u})]$ conditioned on the knowledge of the object spectrum $[O(\vec{u})]$ are also Gaussian. This permits the calculation of the conditional image information entropy $\mathcal{H}_{i|o}$ over the complete system pass-band, $|\vec{u}| < D/(\lambda f)$, as well as the image self information $\mathcal{H}_i$ – and consequently the MI – in terms of the so-far unknown object spectrum statistics. Under the constraints of normalization of probabilities, fixed mean emissive power F of objects over their ensemble, and fixed mean integrated power spectrum G over the system pass-band, we may maximize the MI throughput of the system by varying the statistical ensemble of objects. This maximization has been carried out with the help of Lagrange multipliers, and leads to the following expressions for the maximum statistical information transmitted by the system, or its information capacity (details are in Refs. 7 and 8):

$$\mathcal{C} - \mathcal{C}_\infty \approx -\log\det g, \qquad g_{ij} = \exp[-(1/2)D_\phi(\lambda f|\vec{u}_i - \vec{u}_j|)], \tag{2a}$$

when additive noise is negligible, and

$$\mathcal{C} = \log\left(\frac{e^{1/2}F}{\sqrt{\pi W_E/\Delta A_u}}\right) + \sum_{i=1}^{P}\log\left(\frac{\frac{G}{P} + \frac{1}{P}\sum_j W_E|H_j|^{-2}}{W_E|H_i|^{-2}}\right), \tag{2b}$$

when the additive noise dominates and turbulence is absent. In Eq. (1a), $\mathcal{C}_\infty$ represents the capacity under infinitely strong turbulence for which only the dc pixel carries any information. The symbol P represents the number of spatial frequency pixels, each of area ΔA_u, within the system pass-band, $W_E/\Delta A_u$ the additive noise at each pixel, and H_i is the value of the (noise-free) OTF at the ith pixel. The detailed behavior of the preceding expressions for capacity is discussed in Refs. 7 and 8.

IV. REPORT ON PROGRESS MADE IN THE FINAL PERFORMANCE PERIOD 9/99–8/00 AND FUTURE DIRECTIONS

Two key advances were made during the last performance period. The first involved completion of the paper on the information capacity of a seeing and additive noise limited astronomical imaging system. The paper was submitted, refereed, and published [8] in Optics Communications in April 2000. Secondly, we have applied the notion of Fisher information, a measure of information in statistical data when estimating a particular observable, to gain a solid understanding of the inverse relation between noise and information. Investigating the deep connections between the Shannon and Fisher information measures, as yet largely unexplored in the literature, still lies ahead of us, but our computations of the Fisher information under a variety of imaging conditions have already given us useful insight into these connections. We now describe our, as yet unpublished, studies on Fisher information.

A. Fisher Information and Minimum Estimator Variance

Consider a set of parameters $\{\theta_i\}$ that we estimate by making measurements $\{x_j\}$ that are typically statistical in character. Let the statistics of the measurements be described the probability density function (pdf) $P(\vec{x}|\{\theta\})$. The Fisher information matrix $\mathbf{J}$ is defined by specifying its ij element as

$$J_{ij}(\{\theta\}) = \left\langle \frac{\partial \ln P(\vec{x}|\{\theta\})}{\theta_i} \frac{\partial \ln P(\vec{x}|\{\theta\})}{\partial \theta_j^*} \right\rangle, \tag{1}$$

where the triangular brackets denote averaging over the pdf P itself.

The parameters θ_i are inferred from the data by means of estimators, suitably constructed functions of the data, among which unbiased estimators occupy a special place. For, their averages are defined to be identical, by construction, to the parameters themselves. If $\mathbf{C}$ represents the covariance matrix of any unbiased estimators, $\hat{\theta}_i(\vec{x})$ for the ith parameter θ_i, then $\mathbf{C}$ and $\mathbf{J}$ obey the following Cramer-Rao matrix inequality,

$$\mathbf{C} \geq \mathbf{J}^{-1}, \tag{2}$$

in the sense that $\mathbf{C} - \mathbf{J}^{-1}$ is a nonnegative definite matrix. A necessary condition of this nonnegativity is that the variance of the ith estimator $\hat{\theta}_i(\vec{x})$ cannot be smaller than the ith diagonal element of the inverse Fisher information matrix $\mathbf{J}^{-1}$:

$$\mathrm{var}(\hat{\theta}_i) \geq (\mathbf{J}^{-1})_{ii}. \tag{3}$$

The inverse of the Fisher information thus provides a measure of the minimum uncertainty in the determination of unbiased estimators. It is in this statistical estimation sense that Fisher information represents information. For estimators that are biased, a similar result as Eq. (3) holds.

The imaging problem can be regarded as an estimation problem in which statistical image data are employed to infer the object being viewed. This approach has been employed in the last few years to analyze certain aspects of image formation, detection, and processing, but its relation to Shannon information is not well understood. It is this relation, which we must elucidate before we can hope to clarify the connection between noise reduction and information gain. We have recently turned our attention to this problem.

We describe below our results for the class of estimators in which image data at a single spatial frequency are used to estimate the object spectrum at that frequency. The Fisher information matrix reduces, at each spatial frequency, to a single number that can be easily computed under certain approximations.

For $D/r_0 >> 1$ and in the mid-frequency range, the OTF has nearly Gaussian statistics, and so are the conditional image statistics, given an object, statistics that determine Fisher information:

$$P(I_j|O_j) = \frac{1}{\pi C_{jj}} \exp\left[-(I_j^* - O_j^*\langle H_j\rangle)(I_j - O_j\langle H_j\rangle)/C_{jj}\right], \tag{4}$$

where $\langle H_j \rangle$ is the mean OTF at spatial frequency $\vec{u}_j$ and the image amplitude spectrum variance at that frequency is

$$C_{jj} = |O_j|^2 \sigma_h^2(\vec{u}_j) + \sigma_N^2. \tag{5}$$

The first term on the right hand side of Eq. (5) represents the fluctuation in image data due to atmospheric turbulence and other OTF fluctuations, their variance having been denoted by σ_h^2, while the second term represents the additive zero-mean detector noise taken to be independent of spatial frequency.

For the Gaussian pdf (4), the computation of Fisher information is straightforward. For the case that the object spectral amplitude O_j is purely real, we have for the reciprocal of Fisher information

$$\begin{aligned} J_{jj}^{-1} &= \left\langle \left[\frac{\partial \ln P(I_j|O_j)}{\partial O_j} \right]^2 \right\rangle \\ &= \frac{O_j^2}{4} \frac{[1 + 1/(SNR \cdot \sigma_h^2)]^4}{[1 + 1/(SNR \cdot \sigma_h^2)^2]}, \end{aligned} \tag{6}$$

where SNR is the power-spectrum "signal to noise ratio" at the spatial frequency $\vec{u}_j$,

$$SNR = \frac{O_j^2}{\sigma_N^2}. \tag{7}$$

The OTF variance σ_h^2 under the conditions we have assumed, $D/r_0 >> 1$ in the mid-frequency range, is nearly the same as the speckle tranfer function, and takes the value

$$\sigma_h^2 \approx 0.435 * (r_0/D)^2 * H_0(u_j), \tag{8}$$

where $H_0(u_j)$ denotes the OTF of the telescope in the absence of turbulence, i.e., the diffraction-limited OTF, at frequency u_j.

In Fig. 4, we plot the result (6), which represents the best lower bound on the variance of an unbiased estimator of the object spectrum at that frequency. It is clear that for each D/r_0 value, as SNR increases, the minimum estimator variance decreases as expected. For the same SNR, on the other hand, the minimum estimator variance increases with D/r_0, once again as expected.

Against the best theoretical performance (6) predicted by the Fisher information analysis, we may compare the actual variances of certain well known estimators. We considered three estimators, namely the inverse filter, the Wiener filter, and the filtered back-projection estimator. Only the first of the three is an unbiased estimator, while the third is so only when SNR is infinitely large ($\sigma_N = 0$), but a simple multiplicative correction to Fisher information is all that is needed to include a nonzero bias when computing the theoretical minimum bound on estimator variance.

In the next figure, Fig. 5, we present a plot of the minimum estimator variance for the case of atmospheric turbulence alone, as a function of spatial frequency. With increasing D/r_0, the minimum variance rapidly approaches the value $|O_j|^2/4$, given by Eq. (6) under zero additive noise ($SNR = \infty$).

B. Future Studies

There are several directions in which Fisher-information-based studies of the performance of an imaging system are being conducted by us at present. Among them, the three most notable directions are the following:

1. The connection between Fisher and Shannon information must be fully understood before noise-based and information based limits of performance can be meaningfully compared. There is a slight conceptual difficulty in finding this connection – Fisher information governs the precision of statistical measurements of one or more fixed parameters, while Shannon information applies when statistical distributions of parameters are involved. By extending the definition of Fisher information to statistically specified parameters, the two measures of information can both be made to apply to the imaging problem. This is the approach we adopt. Our early results are encouraging, but thorough simulations are needed for complete understanding.

2. The dynamics of Fisher information under application of prior knowledge, e.g., support, are well worth investigating. Such studies will enable us to understand how prior knowledge of a certain kind affects the distribution of information/noise in the physical and spatial-frequency planes. Like our Shannon information based work, simulations are needed to understand these dynamics well.

3. Iterative image processing techniques like Lucy Richardson and expectation maximization (EM), and maximum a posteriori probability (MAP) methods all suffer from the quiet indignity of having no robust stopping criteria. These techniques, if left to converge, lead to meaningless data-independent reconstructions. Both Shannon and Fisher measures of information have the promise of furnishing objective stopping criteria to decide at what iteration number to regard the reconstruction as having yielded the "best" image. The best image must maximize information contained in the measured data and prior knowledge of the object attributes (like support, positivity, etc).

REFERENCES

[1] P. B. Fellgett and E. H. Linfoot, "On the assessment of optical images," Philos. Trans. Royal Soc. London **247**, 46 (1955).

[2] C. Matson, "Variance reduction in Fourier spectra and their corresponding images

with the help of support constraints," J. Opt. Soc. Am. **A 11**, 97 (1994).

[3] S. Prasad, "Dynamics of Turbulence-Induced Noise in Image Deconvolution with Support Constraint," J. Opt. Soc. Am. A **16**, pp. 1769-1778 (1999).

[4] D. Tyler and C. Matson, "Reduction of Nonstationary Noise in Telescope Imagery Using a Support Constraint," Optics Express **1**, 347-354 (1997).

[5] S. Prasad and D. Tyler, "Information Dynamics in Constrained Image Deconvolution," presented at the *Workshop on Fundamental Issues in Image Formation, Detection, and Processing*, U. of New Mexico, Albuquerque, NM, February 6-7, 1999.

[6] See, e.g., J. Goodman, *Statistical Optics* (Wiley, New York, 1985), Ch. 8.

[7] S. Prasad, "An Information Theoretic Perspective on the Formation,Detection, and Processing of Seeing-Limited Images," *Proceedings of the AMOS Technical Conference, Maui, Hawaii, August 30-September 3, 1999*, pp. 339-349 (1999).

[8] S. Prasad, "Information Capacity of a Seeing-Limited Imaging System," Optics Communications **177**, 119 (2000).

V. Publications Supported by the Contract

[1] S. Prasad, "Dynamics of Turbulence-Induced Noise in Image Deconvolution with Support Constraint," J. Opt. Soc. Am. A **16**, 1769 (1999).

[2] S. Prasad and D. Tyler, "Information Dynamics in Constrained Image Deconvolution," in *Selected Reprints and Summaries, Workshop on Fundamental Issues in Image Formation, Detection, and Processing, U. New Mexico, Albuquerque, February 6-7, 1999.*

[3] S. Prasad, "An Information Theoretic Perspective on the Formation,Detection, and Processing of Seeing-Limited Images," *Proceedings of the AMOS Technical Conference, Maui, Hawaii, August 30-September 3, 1999*, pp. 339-349 (1999).

[4] S. Prasad, "Information Capacity of a Seeing-Limited Imaging System," Optics Communications **177**, 119 (2000).

[5] D. Tyler and S. Prasad, "Noise Transport and Removal in a Speckle Imaging Algorithm," presented at the Annual Optical Society of America Meeting in Santa Clara, September 26-30, 1999.

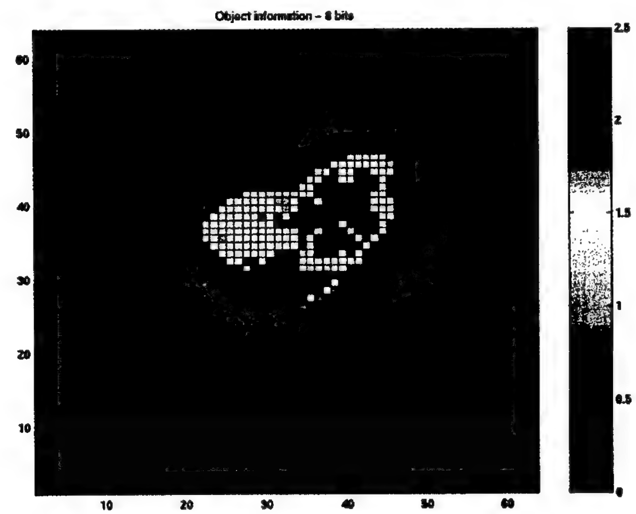

Fig. 1. Distribution of Shannon information in the object set

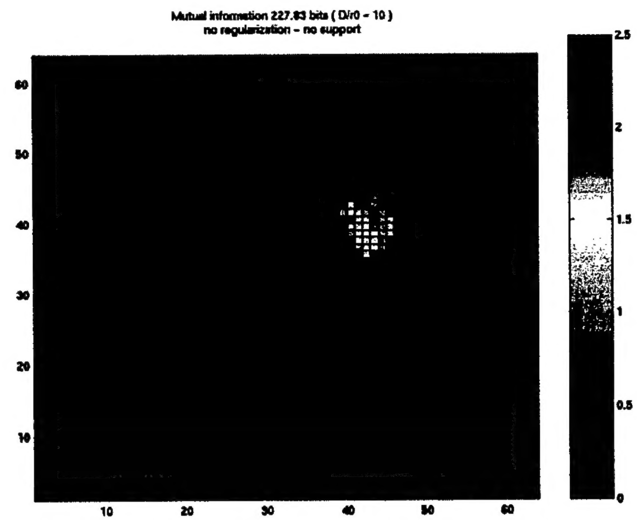

Fig. 2. Mutual information throughput at high turbulence, D/r0=10

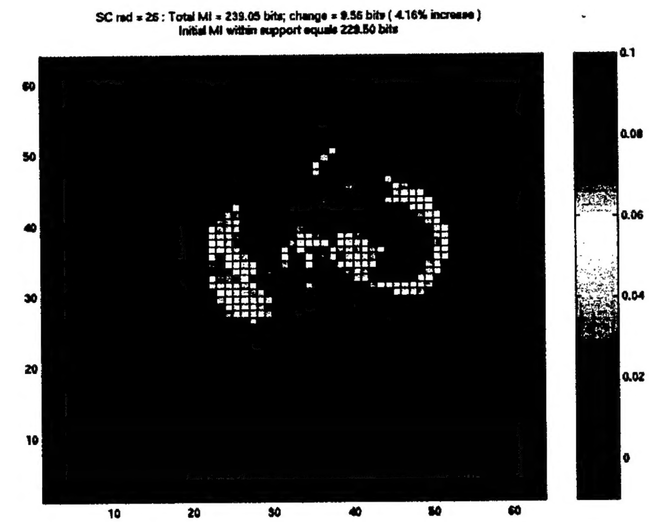

Fig. 3. Improvement of mutual information under application of support

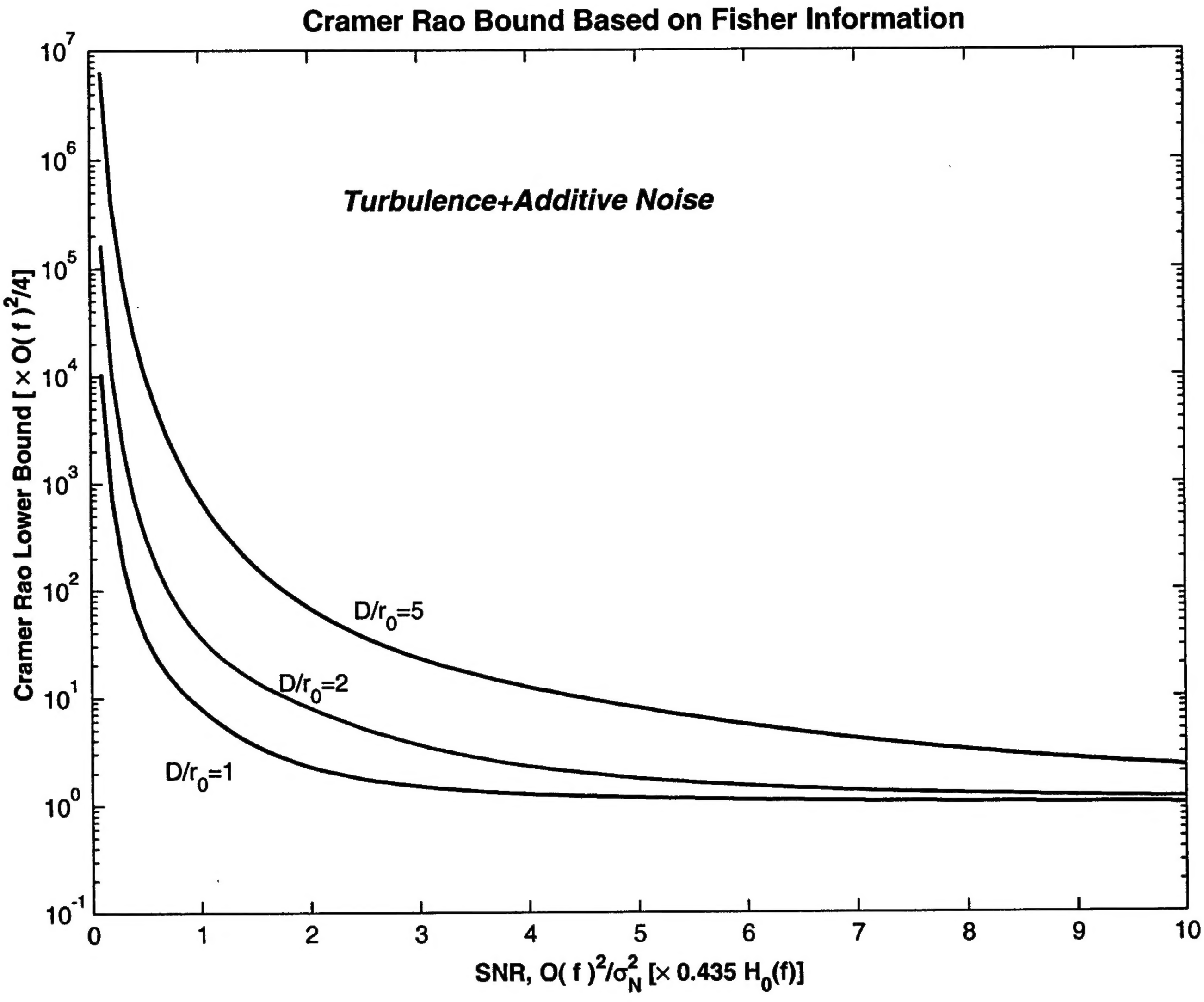
Cramer Rao Bound Based on Fisher Information
Turbulence+Additive Noise
Cramer Rao Lower Bound [× O(f)²/4]
SNR, O(f)²/σ²N [× 0.435 H0(f)]
D/r0=5
D/r0=2
D/r0=1
10⁷
10⁶
10⁵
10⁴
10³
10²
10¹
10⁰
10⁻¹
0
1
2
3
4
5
6
7
8
9
10

Fig. 4

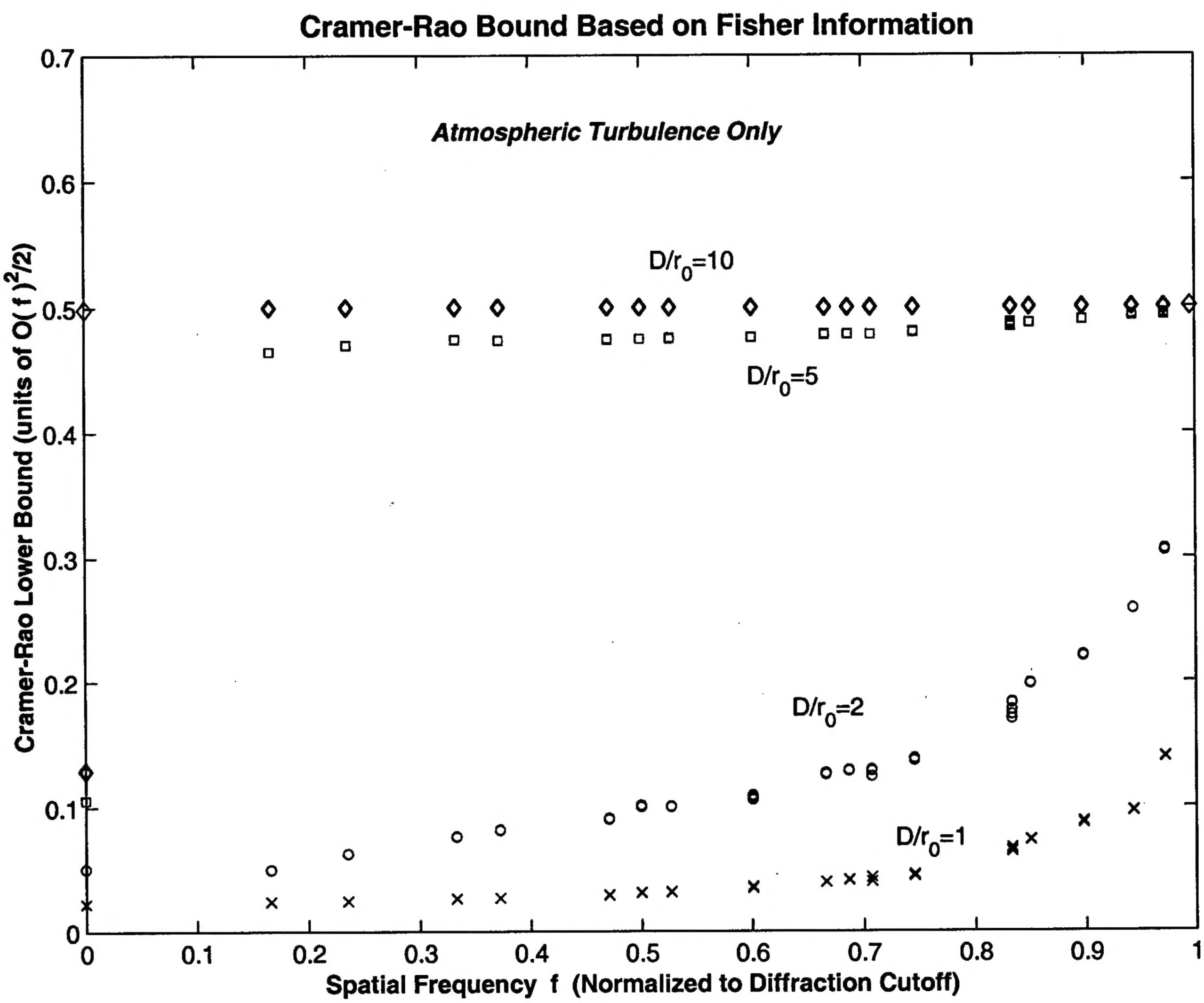
Cramer-Rao Bound Based on Fisher Information
Atmospheric Turbulence Only
D/r0=10
D/r0=5
D/r0=2
D/r0=1
Cramer-Rao Lower Bound (units of O(f)^2/2)
Spatial Frequency f (Normalized to Diffraction Cutoff)
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Fig. 5